Physica A 395 (2014) 200–208

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Asymptotic form of the Kullback–Leibler divergence for multivariate asymmetric heavy-tailed distributions

Javier E. Contreras-Reyes *

División de Investigación Pesquera, Instituto de Fomento Pesquero, Blanco 839, Valparaíso, Chile

HIGHLIGHTS

- We present the asymptotic Kullback–Leibler divergence for multivariate skew-t distributions.
- This model manipulates the skewness and heavy-tail presence of the data.
- The skew-t distribution contains the Student-t, skew-normal and normal distributions.
- Numerical examples show the good performance of our approximation.

ARTICLE INFO

Article history:
Received 8 June 2013
Received in revised form 24 September 2013
Available online 26 October 2013

Keywords:
Kullback–Leibler divergence
Skew-t
Skew-normal
Heavy tails
Skewness

ABSTRACT

An asymptotic expression for the Kullback–Leibler (KL) divergence measure of multivariate skew-t distributions (MST) is derived. This novel class of flexible family distributions incorporates a shape and degree of freedom parameters, in order to manipulate the skewness and heavy-tail presence of the data, respectively. The quadratic form expressions of MST models are used to provide asymptotic measures. Additional inequalities for MST entropy and simulation studies are reported. Finally, the expected values of the KL divergence of a sample correlation matrix obtained by Pearson's correlation coefficient are discussed.

1. Introduction

Shannon entropy [1] and Kullback–Leibler (KL) [2] divergence are perhaps the two most fundamental measures used in information theory and its applications. Frequently, several authors have computed these measures for a large list of distributions, on univariate and multivariate cases [3]. In this sense, [4] provides the properties of Shannon entropy and KL divergence measures for multivariate normal (MN) distribution, among others. Zografos and Nadarajah [3] give the entropy for the multivariate Student-t (MT) distribution and for several other distributions, where this last class of models describes the tail behavior of the marginal distribution.

KL divergence is considered a good indicator of the correlation degree between two finite sets of data, especially if they are affected by noise produced by the interaction of the system, or measurement error [5]. Recent studies considered expected values of KL divergence between a model with a theoretical correlation matrix and a model with a sample correlation matrix, obtained by Pearson's estimator. For example, [5] considered the MN distribution, [6] has extended the above results for the KL entropy to a general class of elliptic distributions and provides the KL divergence between two MT distributions with the same degree of freedom parameters and other asymptotic special cases and, [7,8] concludes that the KL divergence is very good for comparing correlation matrices to an application in financial markets (see [9] for discussion of this issue).

* Tel.: +56 032 2151 513; fax: +56 032 2151 645.
E-mail address: jecontrr@uc.cl.

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.10.035

In the last three decades, non-normal distributions have received substantial interest within scientific literature, especially in regards to skew-elliptical distributions. This class of flexible distribution is characterized for account of skewness and heavy tails as extra parameters with respect to elliptical distributions such as normal and Student-t [10,11], behind the multivariate skew-normal (MSN, [12,13]) and multivariate skew-t (MST, [14,15]) distributions. They have been successfully applied to numerous datasets from a wide range of fields including biological sciences, geophysics, astronomy, engineering and economics. Some recent applications of MST models include those by [11,16–20].

More recent studies have dealt with the calculus of Shannon entropy and mutual information for skew-elliptical distributions [21–26,19,29]. In this sense, [22] addresses the variational inference of skew-normal distributions using the entropy measure as an approximation. Challis and Barber [24] consider the KL divergence minimization problem between a given target density and an approximated variational density, where the skew-normal distribution is considered as a special case. Contreras-Reyes and Arellano-Valle [23] provide explicit formulas of cross-entropy (CE) and KL divergence measures for MSN distributions and, in particular, the Jeffreys divergence measure is used to compare the MN distribution with the MSN distribution, showing that this is equivalent to comparing its respective univariate distributions. Arellano-Valle et al. [19] present entropy and mutual information indexes of multivariate skew-elliptical distributions, where MSN and MST are studied as special cases.

This work provides an approximated KL divergence measure for the flexible family of MST distributions as a possible generalization of KL divergence for MSN distributions given by [23]. Given that the calculus of closed KL divergence expressions for Student-t and skew-t models are complicated, traditional identities and quadratic form expressions given by [27], are used to provide asymptotic measures. Additionally, some inequalities for MST entropy are included and a study simulation is undertaken. Finally, considering the sample correlation matrix obtained by Pearson's correlation coefficient, the expected values of the KL divergences for MSN and MST distributions are computed.

Consider a probability density function (pdf) $f_{\mathbf{Z}}$ associated with a random variable $\mathbf{Z} \in \mathbb{R}^k$. Denote by $H(\mathbf{Z})$ the well-known Shannon entropy introduced by [1]. It is defined by

$$H(\mathbf{Z}) = -\int f_{\mathbf{Z}}(\mathbf{z}) \log f_{\mathbf{Z}}(\mathbf{z}) \mathrm{d}\mathbf{z}. \tag{1}$$

This measure is the expected value of $g(\mathbf{Z}) = -\log f_{\mathbf{Z}}(\mathbf{Z})$, which satisfies $g(\mathbf{1}) = 0$ and $g(\mathbf{0}) = \infty$. Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^k$ be two continuous random variables, the cross-entropy (CE) is defined as follows

$$CH(\mathbf{X}, \mathbf{Y}) = -\int f_{\mathbf{X}}(\mathbf{x}) \log f_{\mathbf{Y}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{2}$$

where it is clear from (2) that $CH(\mathbf{X}, \mathbf{X}) = H(\mathbf{X})$. However, $CH(\mathbf{X}, \mathbf{Y}) \neq CH(\mathbf{Y}, \mathbf{X})$ at least for $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$, i.e., $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution.

Related to the entropy and CE concepts, we can also find divergence measures between the distributions of $\mathbf{X}$ and $\mathbf{Y}$. The most well-known of these measures is the so called Kullback–Leibler (KL) divergence [2],

$$KL(\mathbf{X}, \mathbf{Y}) = \int f_{\mathbf{X}}(\mathbf{x}) \log \left\{ \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{x})} \right\} \mathrm{d}\mathbf{x} \tag{3}$$

which measures the divergence of $f_{\mathbf{Y}}$ from $f_{\mathbf{X}}$ and where the expectation is defined with respect to the pdf $f_{\mathbf{X}}(\mathbf{x})$ of the random vector $\mathbf{X}$. This measure is always positive and is equal to zero if and only if $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$. Therefore, it is not symmetrical because it does not verify the triangular inequality (see, e.g., Ref. [23]). We note that (3) comes from (2) as

$$KL(\mathbf{X}, \mathbf{Y}) = CH(\mathbf{X}, \mathbf{Y}) - H(\mathbf{X}), \tag{4}$$

and $CH(\mathbf{X}, \mathbf{Y}) \geq H(\mathbf{X})$ with equality if only if $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$. For more details about entropy, CE, and KL divergence, see, e.g., Refs [4] and [28].

2. Multivariate skew-elliptical distributions

This novel class of family distributions proposed by [10] is a generalization of several distributions that account for skewness and heavy tails as extra parameters with respect to elliptical distributions such as normal and Student-t. The density function of multivariate skew-elliptical distribution $\mathbf{Z} \sim SE_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta}, h^{(k+1)})$ is

$$p_{\mathbf{Z}}(\mathbf{z}) = 2 f_k(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(k)}) F_1(\boldsymbol{\eta}^\top (\mathbf{z} - \boldsymbol{\xi}); h_s^{(1)}), \tag{5}$$

for $\mathbf{z} \in \mathbb{R}^k$, where $f_k(\mathbf{z}; \boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(k)}) = |\boldsymbol{\Omega}|^{1/2} h^{(k)}(s)$ with $s = \mathbf{z}_0^\top \mathbf{z}_0$, density generator function $h^{(k)}(s)$, $s > 0$, and, $F_1(x; h_s^{(1)})$ is the univariate cdf of density generator function $h_s^{(1)}(u) = h^{(k+1)}(s+u)/h^{(k)}(s)$. The multivariate skew-normal distribution, namely $\mathbf{Z} \sim SN_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\eta})$, is a particular member of the skew-elliptical family with density generator function